

<https://helda.helsinki.fi>

Computational-experimental approach to drug-target interaction mapping : A case study on kinase inhibitors

Cichonska, Anna

2017-08-07

Cichonska , A , Ravikumar , B , Parri , E , Timonen , S , Pahikkala , T , Airola , A , Wennerberg , K , Rousu , J & Aittokallio , T 2017 , ' Computational-experimental approach to drug-target interaction mapping : A case study on kinase inhibitors ' , PLoS Computational Biology , vol. 13 , no. 8 , 1005678 . <https://doi.org/10.1371/journal.pcbi.1005678>

<http://hdl.handle.net/10138/214423>

<https://doi.org/10.1371/journal.pcbi.1005678>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

RESEARCH ARTICLE

Computational-experimental approach to drug-target interaction mapping: A case study on kinase inhibitors

Anna Cichonska^{1,2*}, Balaguru Ravikumar², Elina Parri², Sanna Timonen², Tapio Pahikkala³, Antti Airola³, Krister Wennerberg², Juho Rousu¹, Tero Aittokallio^{2,4*}

1 Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Espoo, Finland, **2** Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland, **3** Department of Information Technology, University of Turku, Turku, Finland, **4** Department of Mathematics and Statistics, University of Turku, Turku, Finland

* anna.cichonska@helsinki.fi (AC); tero.aittokallio@helsinki.fi (TA)



OPEN ACCESS

Citation: Cichonska A, Ravikumar B, Parri E, Timonen S, Pahikkala T, Airola A, et al. (2017) Computational-experimental approach to drug-target interaction mapping: A case study on kinase inhibitors. *PLoS Comput Biol* 13(8): e1005678. <https://doi.org/10.1371/journal.pcbi.1005678>

Editor: Avner Schlessinger, Icahn School of Medicine at Mount Sinai, UNITED STATES

Received: February 3, 2017

Accepted: July 11, 2017

Published: August 7, 2017

Copyright: © 2017 Cichonska et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was financially supported by the Helsinki Doctoral Education Network in Information and Communications Technology HIIT to AC, Academy of Finland [295496 to JR, 269862, 272437, 295504 and 310507 to TA, 272577, 277293 to KW, 289903 to AA], Cancer Society of Finland to KW and TA, Sigrid Jusélius foundation to KW, and the Biocentrum Helsinki

Abstract

Due to relatively high costs and labor required for experimental profiling of the full target space of chemical compounds, various machine learning models have been proposed as cost-effective means to advance this process in terms of predicting the most potent compound-target interactions for subsequent verification. However, most of the model predictions lack direct experimental validation in the laboratory, making their practical benefits for drug discovery or repurposing applications largely unknown. Here, we therefore introduce and carefully test a systematic computational-experimental framework for the prediction and pre-clinical verification of drug-target interactions using a well-established kernel-based regression algorithm as the prediction model. To evaluate its performance, we first predicted unmeasured binding affinities in a large-scale kinase inhibitor profiling study, and then experimentally tested 100 compound-kinase pairs. The relatively high correlation of 0.77 ($p < 0.0001$) between the predicted and measured bioactivities supports the potential of the model for filling the experimental gaps in existing compound-target interaction maps. Further, we subjected the model to a more challenging task of predicting target interactions for such a new candidate drug compound that lacks prior binding profile information. As a specific case study, we used tivozanib, an investigational VEGF receptor inhibitor with currently unknown off-target profile. Among 7 kinases with high predicted affinity, we experimentally validated 4 new off-targets of tivozanib, namely the Src-family kinases *FRK* and *FYN A*, the non-receptor tyrosine kinase *ABL 1*, and the serine/threonine kinase *SLK*. Our sub-sequent experimental validation protocol effectively avoids any possible information leakage between the training and validation data, and therefore enables rigorous model validation for practical applications. These results demonstrate that the kernel-based modeling approach offers practical benefits for probing novel insights into the mode of action of investigational compounds, and for the identification of new target selectivities for drug repurposing applications.

connecting scientists grant [794509103 to AC and BR]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Significant efforts have been devoted in recent years to the development of machine learning models to support different stages of drug development process. Given the enormous size of the chemical universe, such models could offer a complementary and cost-effective means to experimental determination of drug-target interactions, toward prioritization of the most potent ones for further verification in the laboratory. In order to demonstrate the benefits of the prediction models in practical application cases, we carefully evaluated the predictive power of a well-established machine learning model in filling the gaps in existing profiling studies and prediction of target interactions for a new drug candidate. As a specific case study, we focused on kinase inhibitors, which form the largest class of new drugs approved for cancer treatment, but are also known to have wide multi-target activities contributing both to their therapeutic and toxic responses. The high agreement observed between the predicted and experimentally-measured drug-target bioactivities under the implemented rigorous validation setup demonstrates the potential of the machine learning approach, not only for filling the gaps in existing drug-target interaction maps, but also toward off-target interaction prediction for investigational drugs, and finding potential new uses for already approved drugs (drug repurposing).

Introduction

Deregulated kinase activity plays a role in many diseases, hence calling for therapeutic compounds that could effectively inhibit specific members of the protein kinome. Although kinase inhibitors form the largest group of new drugs approved for cancer treatment [1], a majority of them are ATP-competitive, and therefore present a highly promiscuous mechanism of action (MoA), due to the high evolutionary conservation of the kinase ATP-binding pockets [2,3]. The polypharmacological interactions contribute both to therapeutic and toxic responses seen in clinically-approved and investigational kinase inhibitors. Thus, improved knowledge of the complex compound-target binding interactions across the full protein kinome, including both on- and off-target effects, is of high clinical relevance for future drug discovery applications.

Recent technological advances in chemoproteomic approaches, such as thermal profiling [4], have enabled efficient determination of kinome-wide compound potency. Several commercial providers are available for preclinical kinase inhibitor testing *in vitro*, including DiscoverX, Millipore and Reaction Biology. Even though the experimental compound-target interaction mapping is critical to characterizing a compound's MoA, computational methods provide a complementary and cost-effective approach with the potential to accelerate the exploration of the enormous size of the chemical universe, estimated to consist of approximately 10^{20} molecules exhibiting good pharmacological properties [5]. The hypothesis is that *in silico* models could provide fast, large-scale and systematic pre-screening of chemical probes, toward prioritization of the most potent interactions for further *in vitro* or *ex vivo* verification in the laboratory [6–10].

In particular, a lot of work has been devoted to compound-based interaction prediction methods, including quantitative structure-activity relationship (QSAR) models, which aim to relate structural properties of the chemical molecules to their bioactivity profiles [11,12]. Another class of machine learning methods, so called target-based methods, focus on evaluating similarities between amino acid sequences or three-dimensional structures of protein

targets [13]. In these supervised learning approaches, models are trained using available bioactivity data, together with either compound or protein information, which allows then predicting either new target interactions for a given drug or new drugs targeting a given protein. Furthermore, such methods typically focus on a limited set of molecules of interest.

As a more recent class of computational modelling approaches, systems-based frameworks take advantage of the information available on both compounds and targets. For instance, Yamanishi *et al.* proposed a supervised machine learning approach for categorizing drug-target pairs as interacting or non-interacting based on an integrated model of chemical and genomic molecular profiles [14]. Since this seminal work, a wide variety of systems-based prediction methods have been developed that utilize various molecular descriptors and learning techniques, including random forest, neural networks and kernel learning [15–32]. Even though such models may hold a great potential, their computational predictions are rarely being directly verified in the laboratory and, consequently, their practical benefits for the drug discovery or repurposing applications remain largely unknown.

Toward testing the practical potential of systems-based machine learning models, we implemented a computational-experimental framework for prediction and verification of compound-target bioactivity profiles (Fig 1). We focused on a regression problem, where the task is to predict the actual binding affinities, instead of the standard bioactivity classification setting that treats molecular interactions as simple on/off relationships. As a prediction model, we applied a well-established kernel-based regularized least squares learning algorithm (KronRLS [33]), because kernels, in addition to offering a computationally efficient means for increasing the power of linear learning algorithms, are particularly well-suited for capturing and learning complex molecular properties for prediction purposes [34,35].

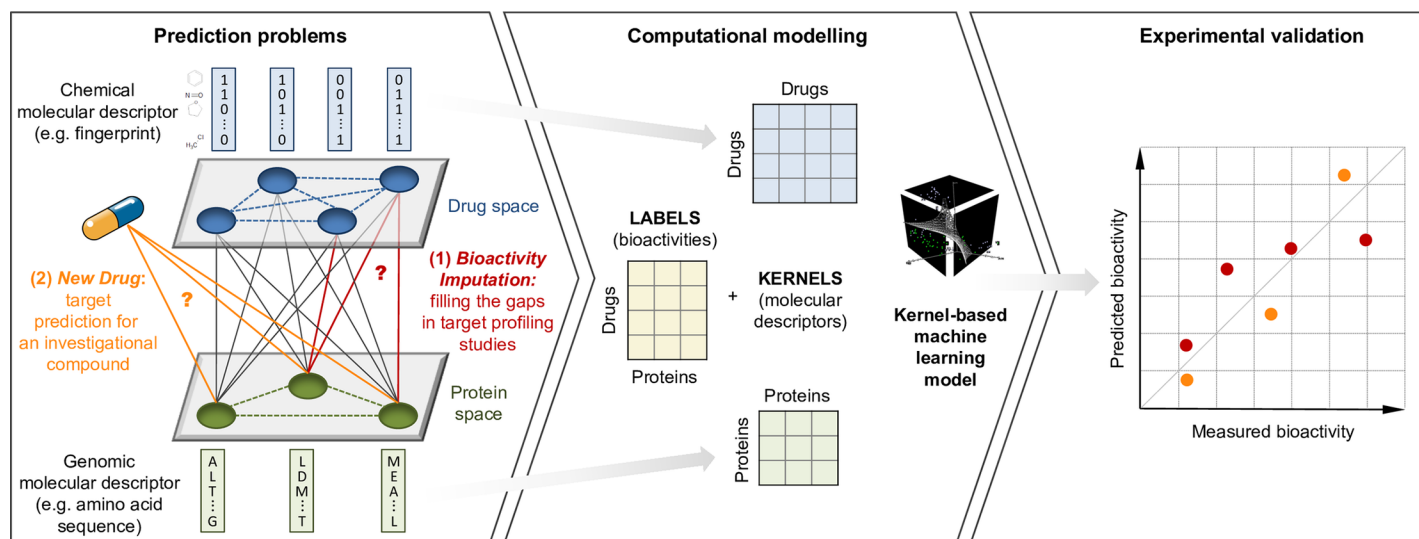


Fig 1. An overview of our computational-experimental framework for prediction and pre-clinical testing of compound-protein bioactivity profiles. Two separate prediction problems are considered: (1) filling the gaps in existing compound-target interaction maps and (2) prediction of target interactions for a new or investigational compound. Molecular descriptors of drug compounds and protein targets are encoded as kernels, and used for binding affinity prediction with a regularized least squares regression model KronRLS. Finally, a subset of predicted compound-protein bioactivities is experimentally tested (see Materials and Methods for details). Since the experimental validations do not exist at the time of making the predictions, this approach effectively assesses any potential model overfitting to the training data only. We chose to use kernel-based models as these are well-suited for representing structured objects, such as molecules, that cannot be accurately described by a standard feature vector. Different types of drug and protein kernels can be calculated using readily available chemical structures and amino acid sequences. The resulting matrices associate all pairs of input objects, and therefore a kernel function can be considered as a similarity measure.

<https://doi.org/10.1371/journal.pcbi.1005678.g001>

The specific contributions of the work are the following. First, we evaluated a large number of molecular descriptors in the form of kernels, including our novel, extended target profile-based protein kernel, and a generic string kernel that has not previously been used in the context of compound-protein interaction prediction. Second, we show how these kernels guided us in filling the gaps in a large-scale compound-kinase interaction map [3]. Third, we experimentally tested a subset of 100 predicted binding affinities, achieving a high correlation of 0.77 between the measured and predicted bioactivities. Finally, we demonstrate the potential of the modelling approach in a more challenging task of predicting target selectivities for such a new candidate compound that has no bioactivity data available for model training. As a specific case study, we used an investigational tyrosine kinase inhibitor tivozanib whose established target profile consists only of 3 on-targets. We experimentally tested and validated 4 out of 7 kinases predicted as tivozanib's off-targets, providing novel insights into its MoA, and thereby extending the potential therapeutic target space of tivozanib.

Results

In the real use cases, the problem of compound-protein interaction prediction should be considered separately under four different scenarios, depending on whether or not the training and test sets share common compounds, proteins, or both (Fig 2 and S1 Table). In the below results sections, we focused on the two most common and practical scenarios of (1) filling the experimental gaps in compound-target profiling datasets (referred to as the *Bioactivity Imputation* scenario, Fig 2A) and (2) prediction of target interactions for an investigational drug compound (referred to as the *New Drug* scenario, Fig 2B). Moreover, S7 Fig shows, for comparison, the results obtained under the symmetric *New Target* scenario (Fig 2C).

Evaluation of molecular descriptors for compound-target interaction inference

A key assumption in the systems-based compound-target interaction prediction algorithm is that similar drug compounds are likely to bind to similar protein targets, and therefore the first challenge lies in the representation and use of molecular similarities in the most predictive way. We encoded here similarities between drugs and similarities between proteins using different types of kernels, constructed based on chemical two- and three-dimensional structures, amino acid sequences, protein structures, and molecular interaction profiles (see [Materials and Methods](#) for details). Such systematic construction of chemical and genomic molecular descriptors resulted in 12 drug kernels and 8 protein kernels. To predict compound-protein binding affinities using the regression setup, we applied a regularized least squares (RLS) model for each pair of drug kernel and protein kernel (KronRLS algorithm, see [Materials and Methods](#)).

For computational evaluation of the predictive performance of various molecular descriptors and optimization of model parameters under separate prediction scenarios, we carried out two systematic nested cross-validation (CV) procedures, i.e., leave-one-out cross-validation (LOO-CV, S12 Fig) and leave-drug-out cross-validation (LDO-CV, S13 Fig), using 16,265 known binding affinities (pK_i values) between 152 kinase inhibitors and 138 protein kinases measured in a large-scale functional bioassay by Metz *et al.* [3] (S9 Fig). We applied LOO-CV to tune model parameters and to evaluate its predictive performance when filling experimental gaps in large-scale target profiling studies (the *Bioactivity Imputation* scenario, Fig 2A), and LDO-CV in the inference of target interactions for a new candidate drug compound (the *New Drug* scenario, Fig 2B). LOO-CV corresponds to the design where scattered missing values are present in otherwise known compound-protein bioactivity matrix, and the aim is to predict

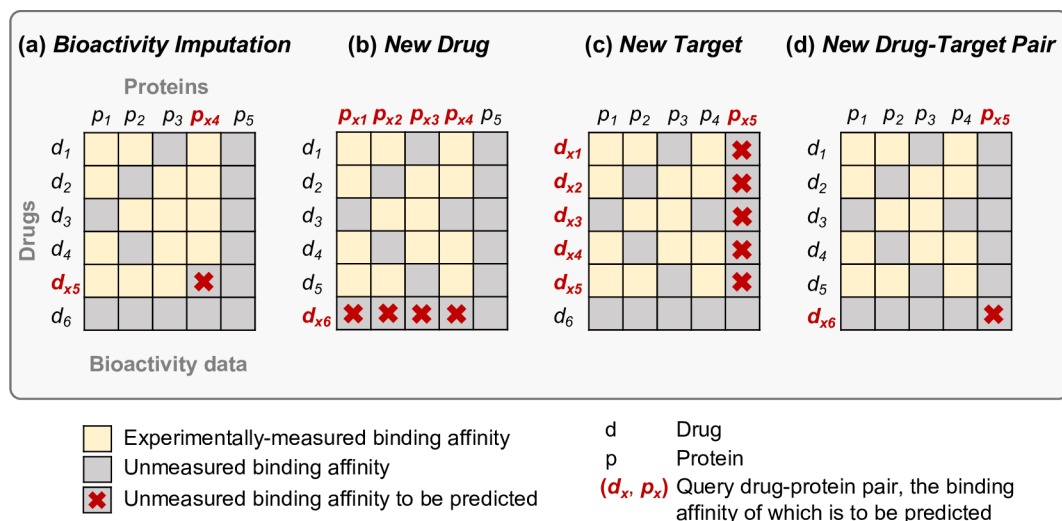


Fig 2. Drug-protein interaction prediction scenarios. (d_x, p_x) denotes a query drug-protein pair, the binding affinity of which is to be predicted. (a) The *Bioactivity Imputation* scenario: both the drug d_x and protein p_x are present in the training set, i.e., there exist known bioactivity values for the drug d_x and protein p_x , but not for their interaction (d_x, p_x). (b) The *New Drug* scenario: the protein p_x is present in the training set, whereas the drug d_x is not, i.e., there exist known bioactivity values for the protein p_x but not for the drug d_x . (c) The *New Target* scenario: the drug d_x is present in the training set, whereas the protein p_x is not, i.e., there exist known bioactivity values for the drug d_x , but not for the protein p_x . (d) The *New Drug-Target Pair* scenario: neither the drug d_x nor protein p_x is present in the training set, i.e., there exist no bioactivity values neither for the drug d_x nor protein p_x . In this work, we focused primarily on two most common and practical prediction scenarios of (a) and (b), which correspond to filling the gaps in existing experimentally-measured drug-target interaction maps and prediction of target interactions for an investigational drug compound, respectively.

<https://doi.org/10.1371/journal.pcbi.1005678.g002>

the missing entries within the training data. LDO-CV, on the other hand, simulates more challenging inference problem, in which the aim is to predict targets of an investigational drug compound, not encountered in the training data (Materials and Methods, Tables 1 and S1).

Bioactivity Imputation: Computational evaluation of filling in the experimental gaps.

In the first task of filling the gaps in experimental bioactivity profiling study, the Gaussian

Table 1. Applied nested cross-validation strategies.

	LOO-CV	LDO-CV ^a
Aim	To evaluate the performance of the prediction model in the task of filling the experimental gaps in existing compound-protein interaction maps (the <i>Bioactivity Imputation</i> scenario, Fig 2A).	To evaluate the performance of the prediction model in the task of predicting target interactions for a new candidate drug compound (the <i>New Drug</i> scenario, Fig 2B).
Outer CV	One entry (compound-protein pair) at a time is removed from the compound-protein interaction matrix \mathbf{Y} , and kept as a test fold.	One row (compound) at a time is removed from the compound-protein interaction matrix \mathbf{Y} , and kept as a test fold.
Inner CV	One entry (compound-protein pair) at a time is removed from the remaining part of the compound-protein interaction matrix \mathbf{Y} , and kept as a test fold.	Five rows (compounds) at a time are removed at random from the remaining part of the compound-protein interaction matrix \mathbf{Y} , and kept as a test fold.

^a In case of LDO-CV, the rows and columns corresponding to the compounds included in the test fold are removed from the drug kernel matrix \mathbf{K}_D before model training.

<https://doi.org/10.1371/journal.pcbi.1005678.t001>

interaction profile drug kernel (KD-GIP) clearly outperformed other compound descriptors (Fig 3A and S1 Fig). This is because, in some cases, even a minor structural difference between chemical molecules causes a striking change in their potency, and such kernel is able to capture this behavior. Among structural fingerprint-based drug kernels, the ones constructed by comparison of two-dimensional substructures defined by PubChem (KD-PubChem-2D) as well as shortest paths between atoms (KD-sp) yielded the best binding affinity predictions. In case of both compounds and proteins, the use of three-dimensional conformations (KD-PubChem-3D, KP-3D-sid, KP-3D-energy) did not lead to improved prediction results.

Among the protein kernels, the protein interaction profile kernel (KP-GIP) and the kernel based on extended target profile built upon Smith-Waterman amino acid sequence comparisons (KP-SW+) showed the best overall performance (Fig 3A). Moreover, KP-SW+, paired with any drug kernel, achieved higher predictive accuracy than its commonly used counterpart KP-SW, which is also based on the Smith-Waterman amino acid sequence alignments but of only proteins included in the training data set, whereas KP-SW+ kernel is calculated based on more comprehensive, global features (see Materials and Methods for details). Notably, generic string kernel worked better with kinase domains (KP-GS-domain) and ATP-binding pockets (KP-GS-atp), compared to full amino acid sequences (KP-GS), indicating their potential for compound-target interaction inference.

Taken together these computational evaluation results under the *Bioactivity Imputation* scenario, the best chemical and genomic molecular descriptor pair in filling the gaps in experimental kinase inhibitor target profiling study was formed by KD-GIP and KP-GS-domain kernels, followed closely by KD-GIP and KP-SW+ kernels, which resulted in high Pearson correlations between the original and predicted compound-kinase binding affinities of 0.829 and 0.828, respectively ($p < 0.0001$, S2 Fig).

New Drug: Computational evaluation of predicting target interactions for new compounds. Predicting bioactivity signatures for a new drug candidate is a lot more challenging task, since such chemical probe has no target interaction data available for the model training. To simulate this setup, we next performed LDO-CV using different pairs of chemical and genomic molecular descriptors, but without including interaction profile kernels. Among drug kernels, the one computed using the shortest paths between atoms (KD-sp) demonstrated the best overall performance (Fig 3B). Together with amino acid sequence-based generic string protein kernel (KP-GS), it formed the most powerful molecular descriptor pair in the inference of interactions for an investigational drug candidate, achieving Pearson correlation of 0.653 ($p < 0.0001$, S2 Fig). We note that a very similar predictive performance was achieved by KD-sp drug kernel coupled either with KP-GS-atp or KP-GS-domain protein kernels (Pearson correlations of 0.651 and 0.649, respectively).

These computational evaluation results demonstrate that an optimal choice of both drug and protein kernels depends on the practical application use case, such as whether one is interested in the *Bioactivity Imputation* or *New Drug* prediction scenarios. Computational CV protocols provided us useful tools for optimizing the prediction models, which is a critical prerequisite for the achievement of high-quality binding affinity predictions, before going into the more laborious and expensive experimental validations.

Filling the experimental gaps in large-scale kinase inhibitor target profiling study

Next, we trained the KronRLS algorithm with 16,265 bioactivities between 152 kinase inhibitors and 138 kinases measured in the study by Metz *et al.* [3], together with the best-performing under the *Bioactivity Imputation* scenario drug interaction profile kernel (KD-GIP) and

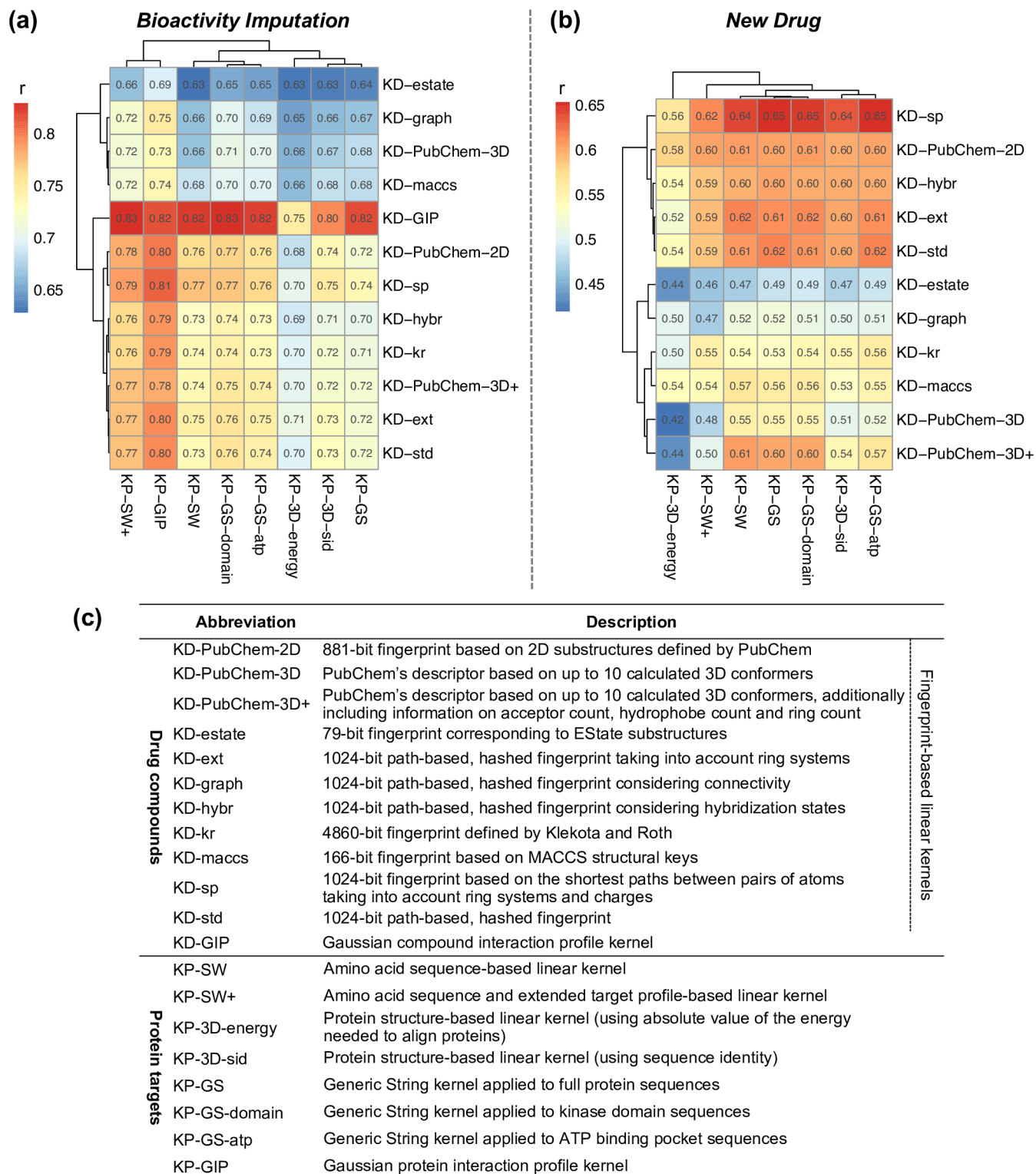


Fig 3. Computational evaluation of the model predictions. (a) Leave-one-out and (b) leave-drug-out cross-validation results. The prediction accuracy was evaluated with Pearson correlation (r) between binding affinities (pK_i) from the study by Metz *et al.* [3] and those predicted using KronRLS algorithm with different pairs of compound (rows) and protein (columns) molecular descriptors encoded as kernel matrices (c). The corresponding root mean squared error (RMSE) values are shown in S1 Fig. Of note, Gaussian interaction profile drug kernel (KD-GIP), which resulted in the highest predictive performance under the *Bioactivity Imputation* scenario (a), was not evaluated under the *New Drug* scenario (b),

because it is constructed based on the bioactivity profile of a drug to be predicted, that is, using information that in practice is unavailable when predicting target interactions for a new investigational drug compound.

<https://doi.org/10.1371/journal.pcbi.1005678.g003>

kinase domain-based generic string protein kernel (KP-GS-domain). We then used the optimized model to predict the remaining 4,711 binding affinities that were missing in this experimentally-measured compound-kinase interaction map (S9 Fig).

To assess the model's practical utility, we experimentally tested a set of 100 predicted binding affinities between 5 drug compounds (cediranib, lapatinib, gefitinib, pazopanib and vx-745) and 20 kinases (*ABL1*, *AXL*, *BRK*, *BTB*, *EGFR*, *FAK*, *FYN A*, *HER2*, *HER4*, *IGF1R*, *InsR*, *ITK*, *JAK3*, *KDR*, *LCK*, *LYN B*, *PYK2*, *SRC*, *SYK*, *TRKA*). Among these, new potential interactions, not present in the Metz *et al.* dataset, were predicted for cediranib, lapatinib and gefitinib (S4 Fig). We note that pazopanib presented very high prediction accuracy, despite of having a sparse binding affinity profile available for the model training. On the other hand, vx-745 had no potent activities, either measured or predicted, against any of the kinases, and therefore it served as a negative control in the validation (S4 Fig). We tested the predicted bioactivities using a cell-free ADP-Glo Kinase Assay (see [Materials and Methods](#) for details).

We observed a relatively high Pearson correlation of 0.774 ($p < 0.0001$) between the model-predicted (pK_i) and experimentally-measured (pIC_{50}) bioactivities among the 100 compound-kinase pairs (Fig 4A). The IC_{50} readout from our assay, similarly to the inhibition constant K_i in the Metz *et al.* study, indicates the concentration of the compound needed to inhibit enzymatic activity of a kinase by 50%. Even though IC_{50} is known to depend on the concentration of the enzyme, inhibitor, and substrate, along with other experimental conditions, whereas K_i is an intrinsic, thermodynamic quantity independent of the substrate [36], recent studies have shown a sufficiently high level of association between pIC_{50} and pK_i readouts, permitting their reliable comparison [37,38]. We also observed a strong technical correlation of 0.769 ($p < 0.0001$) between pIC_{50} readouts from our profiling assay and pK_i values measured in the study by Metz *et al.* (S5 Fig), which supports the feasibility of our experimental validations. Furthermore, based on the published information [39], the ATP concentration used in our assay (10 μ M) is expected to be below, or in some cases equal to, the ATP K_m values of the kinases tested, suggesting that the IC_{50} values should be very close to the respective K_i values.

We also compared both the model-predicted and experimentally-measured interaction mapping to the results from another large-scale binding assay by Davis *et al.* [40]. In this study, 72 clinically relevant kinase inhibitors were profiled against 442 kinases, providing, for each compound-kinase pair tested, dissociation constant K_d , indicating the tendency of a larger molecular complex to dissociate reversibly into the component molecules. We again observed a very good agreement (correlation of 0.796, $p < 0.0001$) between the computationally-predicted (pK_i) and measured (pK_d) binding affinities across the overlapping 95 compound-kinase pairs (S5 Fig). We noted even a higher technical correlation of 0.916 ($p < 0.0001$) between pK_d values from Davis *et al.* study and pIC_{50} values from our experimental assay (S5 Fig). Notably, a comparison between predicted pK_i values and measured pK_d readouts across a larger set of 2,662 compound-kinase pairs overlapping between Metz *et al.* [3] and Davis *et al.* [40] studies resulted in a lower correlation (0.642, $p < 0.0001$, S6 Fig), compared to that when considering only the pairs included in our experimental assay (0.796, $p < 0.0001$, S5 Fig). A high pK_d indicates that a substrate is more likely to be bound to an enzyme, whereas pK_i measures the potency of a drug. Even though both high pK_i and pK_d values are considered as indicators of drug activity, a drug with high pK_i does not necessarily result in a high pK_d (S6 Fig).

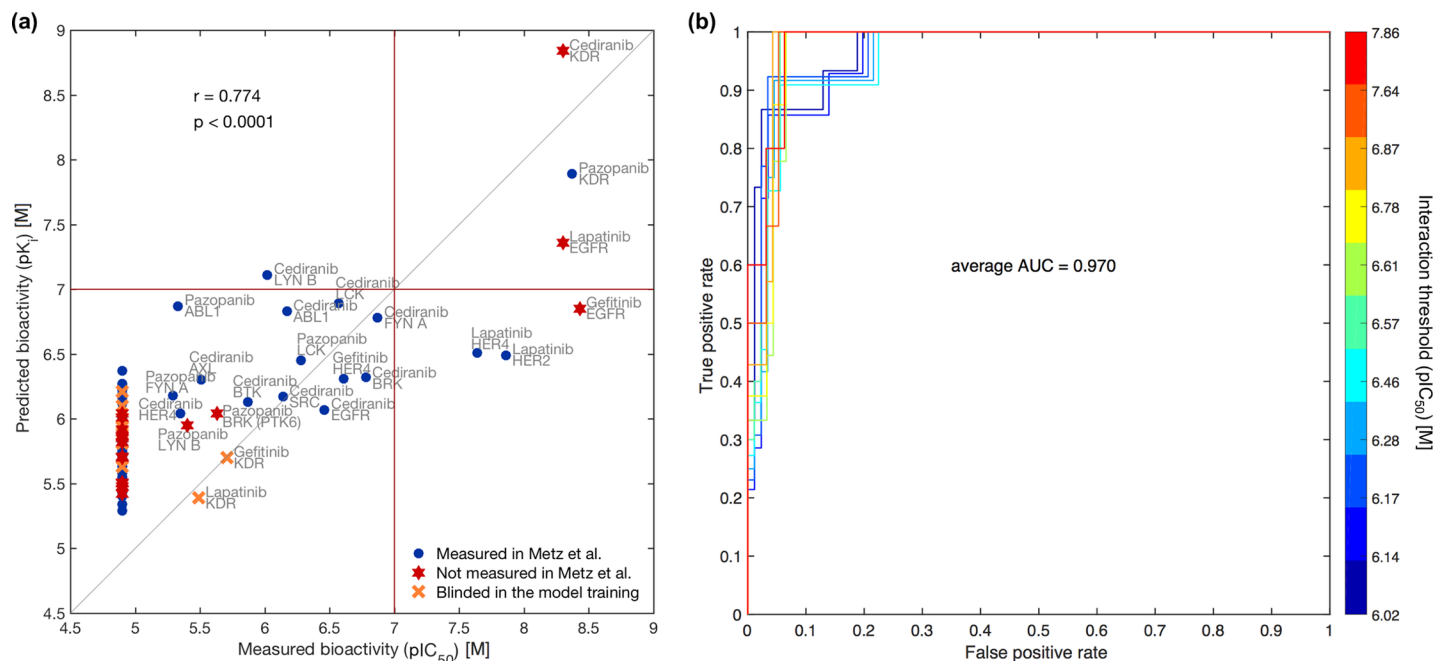


Fig 4. Comparison between computationally-predicted and experimentally-measured bioactivities. (a) Scatter plot between bioactivity values of 100 compound-kinase pairs (detailed in S2 Table). r indicates Pearson correlation. The orange cross points correspond to compound-kinase pairs tested in the study of Metz *et al.* but randomly blinded by us in the training of the model, forming an additional validation set. When no clear interaction between compound and kinase was observed in our experimental assay, the pC_{50} value was set to 4.9 M, corresponding to the highest drug concentration used in our screen (12,500 nM). The higher the pK_i/pC_{50} value, the stronger the affinity between the two molecules. Red lines mark a relatively stringent interaction threshold (7 M), distinguishing the top left corner as the region containing false positive interaction predictions, and the bottom right corner as false negative predictions. (b) A set of receiver operating characteristic (ROC) curves to investigate the model performance as a function of varying activity threshold. We applied 11 different interaction threshold values from the pC_{50} interval [6 M, 8 M] to binarize the experimentally-measured bioactivities into true class labels, and then determined how accurately the model can discriminate between the interacting and non-interacting compound-kinase pairs. The average area under the ROC curves (AUC) equals 0.970.

<https://doi.org/10.1371/journal.pcbi.1005678.g004>

As expected, the correlation between the model-predicted and measured by Metz *et al.* pK_i values in the training data (excluding pairs blinded in the model training, marked with an orange cross points in Fig 4A) was somewhat higher than that for the missing compound-kinase pairs (correlation of 0.802, $p < 0.0001$, S5 Fig). Of note, our experimental assay confirmed computationally-predicted high binding affinities between cediranib-KDR, lapatinib-EGFR and pazopanib-KDR, the two first of which were even not measured in the study of Metz *et al.* Further, although some bioactivities missing in Metz *et al.* dataset corresponded to compound-kinase pairs already tested in other studies (e.g. lapatinib-EGFR measured in the assay by Davis *et al.*), these were not used in the training of our model. Taken together, the observed high agreement between the predicted and experimentally-measured bioactivities demonstrates the potential of the kernel-based modeling framework with appropriately chosen kernels for filling the gaps in existing compound-target interaction maps.

Prediction of off-target interactions for an investigational kinase inhibitor tivozanib

Finally, we tested whether the optimized model can also predict target interactions for a new chemical probe, which has no profiling data available for the model training. We used here tivozanib as an example of an investigational tyrosine kinase inhibitor, known to be potent towards all three VEGF receptors (*FLT1*, *KDR*, *FLT4*) [41]. Beyond VEGFRs, however, the

target profile of tivozanib has otherwise remained poorly characterized, including its potential off-targets. We therefore again used 16,265 binding affinities between 152 compounds and 138 kinases measured in the study by Metz *et al.* to train the KronRLS model with shortest paths between atoms-based drug kernel (KD-sp) and amino acid sequence-based generic string protein kernel (KP-GS), which were found to perform best under the *New Drug* scenario. Since the model should always be tuned separately under distinct prediction scenarios, even if the training dataset is the same, the model used here and the one described in the previous section differ in their chosen kernels and optimized value of the regularization parameter. With the optimized model, we predicted the bioactivity of tivozanib against the set of 138 kinases (S3 Table).

As the first positive control, the model correctly predicted high potency of tivozanib against its known on-targets *FLT1*, *KDR* and *FLT4* (Fig 5A and S3 Table). To further assess the quality of the predictions, we used publicly available bioactivity data from the study of Gao *et al.* who profiled 158 kinase inhibitors, including tivozanib, for their inhibitory activity at 1 μ M and 10 μ M against 234 kinases [42]. Although the concentrations adopted in this screen were too high for pre-clinical testing of positive interactions, we used these data to evaluate the negative predictions from the model. In total, 64 out of 82 kinases with low predicted affinities ($pK_i < 6$ M) were screened by Gao *et al.* Among these, 59 kinases (92%) have at least 50% of the activity remaining at the high compound concentration of 1 μ M (S3 Table), thus effectively validating the model's negative predictions (Fig 5B).

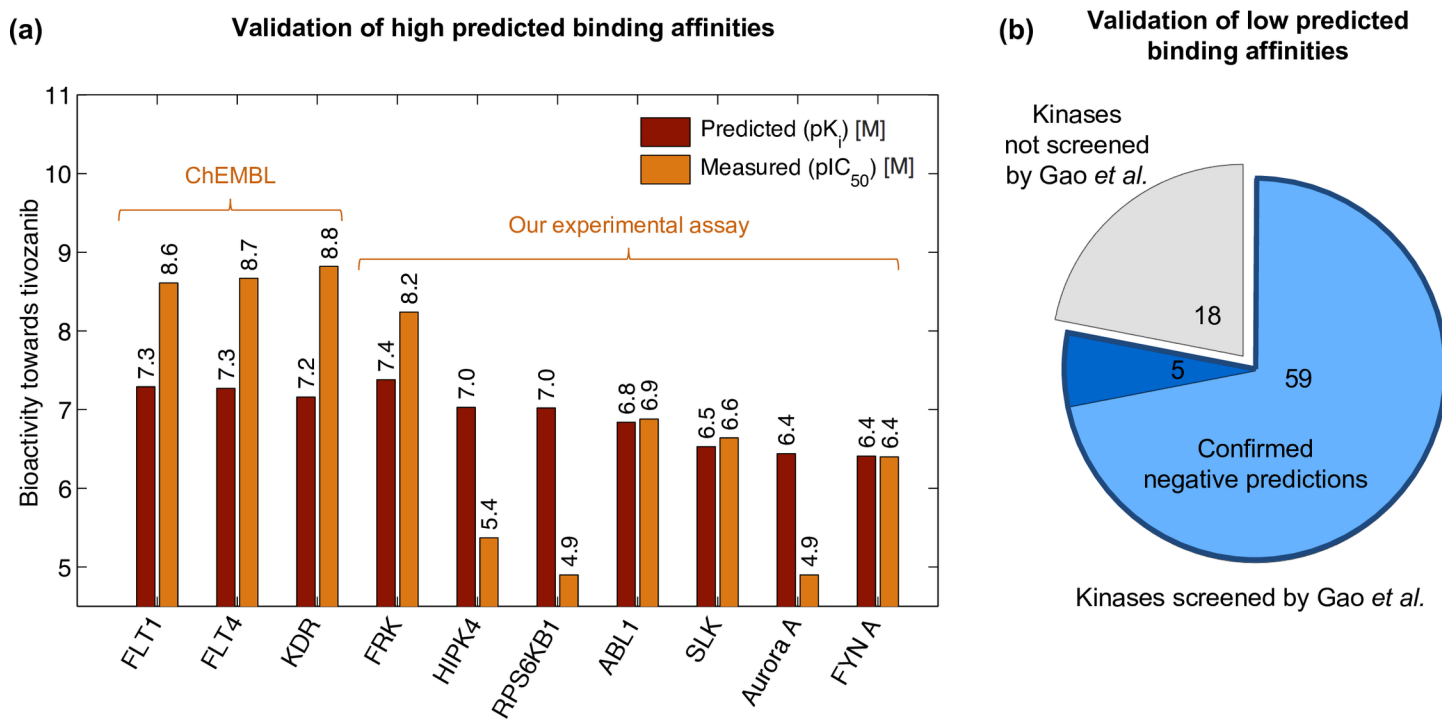


Fig 5. Prediction of target interactions for an investigational kinase inhibitor tivozanib. (a) Predicted and measured bioactivity profiles of tivozanib against its 3 established on-targets (*FLT1*, *FLT4*, *KDR*; average bioactivity from ChEMBL; S3 Table) and 7 predicted off-target kinases tested in our experimental assay. Pearson correlation $r = 0.668$ ($p = 0.035$). When no clear compound-kinase interaction was observed in our assay, the pIC_{50} value was set to 4.9 M, corresponding to the highest drug concentration used (12,500 nM). Predicted values belong to approximately constant range because we focused on experimental validation of the model-predicted off-target interactions. Three of them turned out to be false positives, and therefore the range of experimental results varies more than the range of predicted values. (b) Evaluation of negative interaction predictions from the model. Among 82 kinases with low predicted binding affinities ($pK_i < 6$ M), 64 were screened by Gao *et al.*, and 59 of these are not likely targets of tivozanib (as they have at least 50% of the activity remaining at the high compound concentration of 1 μ M).

<https://doi.org/10.1371/journal.pcbi.1005678.g005>

We next went on and tested experimentally 7 predicted off-target interactions (*ABL1*, *Aurora A*, *FRK*, *FYN A*, *HIPK4*, *RPS6KB1*, *SLK*). These 7 kinases were selected among the set of 25 kinases with the highest predicted binding affinities by focusing on off-targets unique to tivozanib. Specifically, we compared the predicted target interaction profile of tivozanib to other VEGFR inhibitors found in the ChEMBL database [43]. For instance, *RET* was not selected, even though it was predicted to have high potency towards tivozanib ($\text{pIC}_{50} = 6.9 \text{ M}$), since it is targeted by 76% of VEGFR inhibitors in ChEMBL (potency of at most 100 nM), whereas *FYN A* was included in our experimental assay because it is targeted only by 26% of VEGFR inhibitors (S3 Table). We tested the predicted bioactivities using a cell-free ADP-Glo Kinase Assay (see Materials and Methods for details).

Among the pre-selected off-target predictions, our experiments confirmed strong binding affinity between tivozanib and 4 out of the 7 tested kinases (57%), namely *FRK*, *ABL1*, *SLK* and *FYN A* (Fig 5A). Statistical significance of this success rate depends on the underlying distribution of the true target space of tivozanib, which is unknown. However, if one assumes that no more than 18 of 138 considered kinases are actual targets of tivozanib (13%), then the observed overlap is significant ($p < 0.05$, hypergeometric distribution). The observed correlation of 0.668 ($p = 0.035$, Fig 5A) between the predicted and measured binding affinities can be considered relatively high, given the rather limited spectrum of kinases tested and the fact that instead of selecting the top predicted off-targets only, we focused on the set of kinases that were unique to tivozanib among 25 kinases with the highest predicted binding affinities against it. Our experimental results provide not only novel insights into the MoA of tivozanib, but also demonstrate how the *in silico* framework offers a cost-effective tool for prioritizing the most promising target interactions of an investigational compound for further experimental evaluation.

Discussion

Recently, a lot of effort has been placed on the development of systems-based machine learning models that could aid drug discovery process in terms of providing cost-effective compound-target bioactivity predictions. Their main differences lie in the way how the models construct and treat molecular descriptors, and utilize various learning techniques, including those based on random forest [16,19,21], kernel learning [15,22,23,32], recommender systems [26], matrix factorization [20,25,26], Boltzmann machines [27], deep neural networks [16,17], logistic regression [28], learning to rank [29], and ensemble learning [30,31]. Although such models have been shown to perform well in cross-validation setups, their practical benefits still remain largely unknown due to the lack of systematic verification, using targeted experimental assays, carried out sub-sequent to the prediction phase. In model-guided mapping applications, the validation experiments are performed based on the model predictions. Such experimental validation setup effectively avoids any possible information leakage between the training and validation data, since the validation data does not exist at the time of making the predictions. The computational-experimental approach, implemented in this study, therefore makes it impossible to overfit the model to the training data. Here, we used the approach to evaluate the predictive power of a well-established kernel-based learning technique [33]. We chose this model family since kernel regression approaches have proved good performance in recent computational studies, including prediction of drug-target interactions [32], peptide-protein binding affinities [44], drug sensitivities in cancer cell lines [45], as well as in metabolite identification [46] and QSAR modelling [35].

Taking into account that molecular interactions are not simple on/off relationships, we focused on a binding affinity prediction problem, using the RLS regression model with a Kronecker product kernel (KronRLS). Our systematic evaluation of the predictive performance of

various descriptors in the form of kernels revealed that their choice has a critical impact on the prediction accuracy. This is expected because kernel matrix is a central component of the kernel-based learning algorithm as it should capture our prior belief on the relationships between the input objects. In particular, known binding affinities, even if sparse, constitute an important information source not only for model training but also for the kernel matrix construction. Purely structure-based chemical descriptors were not able to fully capture the changes in compounds' activity caused by minor structural differences. Furthermore, we introduced a novel protein kernel (KP-SW+), based on extended target profile, and showed how it consistently outperformed its commonly-used counterpart, in which the Smith-Waterman amino acid sequence alignment is adapted exclusively to proteins included in the dataset of interest (KP-SW). This was evident particularly under the *New Target* setup (Fig 2C), where the aim is to predict compounds targeting a new protein not encountered in the training data (correlation of 0.669 for KP-SW+ vs. 0.506 for KP-SW; S7 Fig). Under the same setting, we also observed a clear advantage of using, for the first time in the context of drug-protein interaction inference, generic string kernel applied to kinase domains and ATP-binding pockets over full protein sequences (correlations of 0.651 for KP-GS-domain, 0.628 for KP-GS-atp, 0.508 for KP-GS; S7 Fig). The majority of kinase inhibitors, including those considered here, bind to ATP-binding pockets, and short sequences of these pockets are included within the kinase domain sequences, thus capturing also the neighbouring context. However, polypharmacological activities of kinase inhibitors, originating from the conservation of kinase ATP-binding pockets, make the prediction problem highly challenging, and better accuracies are likely obtained with compounds having more distinct target profiles. The methodology introduced here could equally well incorporate other compound and protein classes, such as ion channels or G-protein-coupled receptors (GPCRs), but further work will be required to investigate its practical performance under various scenarios using both computational and experimental validations. For instance, it remains an open question which kernels should be calculated to best represent such extended pharmacological spaces.

Our results also demonstrate the importance of a proper evaluation procedure of the *in silico* models. A rigorous computational CV protocol is critical to ensure realistic performance estimates for the optimized models. In particular, the lack of the nested CV strategy in the model selection may lead to over-optimistic prediction results [32]. It is also important that CV design reflects the practical application use case of the model. Given a query drug-protein pair (d_x, p_x) , four different prediction scenarios can be distinguished, depending on whether there exist other compounds with measured bioactivities against p_x , or proteins with measured bioactivities against d_x (Fig 2 and S1 Table). In turn, different types of CV designs need to be implemented in order to tune the model parameters and to evaluate its predictive performance. Here, we focused on the two most common and practical scenarios of the *Bioactivity Imputation* (Fig 2A) and *New Drug* (Fig 2B). Additionally, we provided the CV results under the *New Target* setup (Fig 2C) in S7 Fig. We first adopted LOO-CV, a design that simulates scattered missing values in otherwise known compound-target bioactivity map, to optimize the model and assess its performance in filling the experimental gaps (*Bioactivity Imputation*). Next, LDO-CV was applied in predicting target interactions for new candidate compounds having no measured bioactivity data for the model training (*New Drug*). The latter is much more challenging task, which was also demonstrated in our results; the correlation between the measured and predicted binding affinities under the *Bioactivity Imputation* setup was much higher (0.829) than under the *New Drug* scenario (0.653). However, even in the latter scenario, we still obtained a large number of statistically significant correlation values (S3 Fig). We further observed a high average AUC values under the *Bioactivity Imputation* setup (0.945, S2 Fig), but also under the *New Drug* prediction scenario (0.853, S2 Fig), which indicates that

the model is able to discriminate well the interacting from non-interacting compound-kinase pairs. As expected, the classification accuracy increased with the increasing activity threshold as the true positive set includes a growing number of most likely interactions.

In practical applications, the method requires features extracted from both compounds and proteins, such as readily available chemical two-dimensional structures and amino acid sequences, respectively, based on which kernels can be then calculated. However, if one is interested in a bioactivity of an investigational drug against a protein with unknown sequence, a reasonable prediction accuracy can still be achieved if there exist other compounds with measured bioactivities against a query protein (the *New Drug* setup). The practical solution is to replace the protein kernel matrix with an identity matrix, which implies that each protein will be considered similar to itself only and, effectively, the model will use just known bioactivity data and drug-drug similarities during both training and prediction phase. In particular, we noted only a small drop in Pearson correlation after replacing the protein kernel (KP-GS) with the identity matrix (0.645 vs. 0.653). This observation is expected in multitask or transfer learning problems, such as the *New Drug* or *New Target* setups, where one of the similarities is essential for generalizing to new instances (drug-drug similarities under the *New Drug* setting, protein-protein similarities under the *New Target* scenario).

Ultimately, even though proper CV design is crucial to tune the model and assess its performance, subsequent experimental verification in the laboratory is the only way to really demonstrate the practical utility of the model predictions for drug discovery applications. The relatively good agreement between the computationally-predicted and experimentally-measured bioactivities validated the potential of the kernel-based algorithm, not only for filling the experimental gaps in existing drug-target interaction maps, but also in later stages of the drug development process, including prioritizing new target interactions of investigational compounds for further experimental evaluation, hence assisting in understanding of their MoA. Even though *in silico* inference of target interactions for new candidate drug compounds is a highly challenging task, our results with tivozanib suggest that, given enough-representative and high-quality training data, reliable off-target interaction predictions can be made. In addition to tivozanib, we initially considered also three other investigational kinase inhibitors, namely fedratinib, vx11e and ulixertinib (a compound derived from vx11e). However, we finally selected tivozanib because, unlike for the other compounds, its known on-targets were placed among the strongest predicted target interactions (S3 Table). This indicates that the primary on-target space of tivozanib is well-represented in our training data (S8 Fig). In the future, it is therefore important to profile and build up more diverse training data sets, including more examples of compounds targeting different kinase and other target classes.

Tivozanib was originally-developed as a VEGFR inhibitor meant to block angiogenesis by targeting endothelial cells in the tumor vasculature. However, its MoA has not yet been fully elucidated. Based on the model predictions, our experimental assay confirmed the two Src family kinases *FRK* and *FYN A*, as well as the non-receptor tyrosine kinase *ABL1* and serine/threonine kinase *SLK* as tivozanib's off-targets. Our results highlight that tivozanib has an unusual target spectrum beyond the VEGFR family of kinases, and this suggests that the best anti-cancer use of this compound may not be in diseases where other VEGFR inhibitors with different target profiles have proven effective, but rather in ones where the target spectrum of tivozanib is more unique. For example, it can be hypothesized that tivozanib may have powerful activity in Src-family kinase addicted cancers, where it would target both angiogenesis and the cancer cells directly. Tivozanib has been shown to have a better safety profile than other marketed tyrosine kinase inhibitors and, currently, it is undergoing several clinical trials for the treatment of renal cell carcinoma (NCT03136627), refractory advanced renal cell carcinoma (NCT02627963), metastatic and non-resectable soft tissue sarcomas (NCT01782313),

advanced liver cancer (NCT01835223), recurrent ovarian, fallopian tube, or primary peritoneal cancer (NCT01853644), and advanced prostate cancer (NCT01885949; July 2017). It will be interesting to see which of these trials will report successful treatment outcomes, and which will be terminated due to insufficient efficacy or toxicity.

Although presented here results are promising, there is much room for improvement. For instance, we formulated the predictive model using only one pair of chemical and genomic descriptors at a time. However, even better accuracies could be obtained with a multiple kernel learning framework, which integrates multiple biological and molecular data sources, along with learning their importance for the prediction task [47]. Additional improvement of the predictive performance could be achieved also by creating more sophisticated kernelized molecular descriptors, for instance, by comparing three-dimensional structures of protein binding pockets. Furthermore, we used here as the training data a single yet very comprehensive kinase inhibitor profiling assay containing large number (16,265) of measured compound-kinase binding affinities, spanning different kinase branches (S11 Fig). However, as a future direction, we plan to work on integrating bioactivity values originating from various target profiling experiments and bioactivity end-points into a single model [3,19,40,48]. Recently-initiated community-driven efforts, such as Drug Target Commons (<https://drugtargetcommons.fimm.fi>), which aim to collectively extract, manage and curate high-quality compound-target bioactivity data from public databases, literature and other resources, as well as annotate them with a common ontology, will be essential to facilitate the data standardization and computational modelling purposes. Nevertheless, we hope the current work provides a useful starting point and a practical guide on how to computationally prioritize the most promising target interactions for further experimental evaluation.

Materials and methods

Bioactivity dataset

We used publicly available compound-target interaction map generated by Metz *et al.* using a large-scale functional bioassay, which measured the concentration of a compound needed to inhibit the reaction catalysed by a kinase enzyme of interest by 50% [3]. The readout corresponds to an inhibition constant K_i , typically expressed in the logarithmic scale as $pK_i = -\log_{10}K_i$. Although the universal activity threshold cannot be explicitly defined for each compound-kinase pair, the higher the pK_i value, the stronger the binding affinity between the compound and kinase.

Among molecules included in the screen, 201 compounds are present in ChEMBL [43], and 169 proteins belong to the group of catalytically active human protein kinases [37]. The study is not complete, and therefore we used here a subset of these data: kinases and compounds for which at least 30% of the binding affinity values were measured, resulting in 152 drug compounds and 138 kinase targets. In total, there are 16,265 binding affinities in this selected interaction map (S9 and S10 Figs). Of note, most of the compounds constitute investigational, not yet FDA-approved, chemical probes.

Prediction model

In supervised learning tasks, training data has the form $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where N denotes the number of training examples, $\mathbf{x}_i \in X$ is an input object represented as a vector with the feature values (e.g. a compound represented as a fingerprint vector) and $y_i \in Y$ is its known associated label value (e.g. a potency of a compound against a certain protein). The aim is to find a prediction function f that models the relationship between \mathbf{x}_i 's and y_i 's, and which can then be used

to predict the label values for new instances outside the training space. Classical algorithms search for linear dependencies but often the actual relations underlying the data are highly nonlinear. Kernels offer the advantage of increasing the power of the linear learning machines by providing a computationally efficient way of projecting the input data into a high-dimensional feature space. A linear model in this implicit feature space corresponds to a nonlinear model in the original space. A separation of the statistical learning technique and the data representation is another convenient attribute of kernels.

Formally, a kernel is a function k that for all $\mathbf{x}, \mathbf{z} \in X$ satisfies $k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$, where ϕ is a mapping from the input space X to an inner product high-dimensional feature space F : $\phi: \mathbf{x} \in X \rightarrow \phi(\mathbf{x}) \in F$, and it can be considered as a similarity measure between two objects \mathbf{x} and \mathbf{z} . It is, however, often possible to avoid the explicit computation of the mapping ϕ , and define the kernel directly in terms of the original input data items by replacing the inner product $\langle \cdot, \cdot \rangle$ with an appropriately chosen kernel function satisfying certain mathematical properties (so-called kernel trick). Kernels are particularly handy for calculating similarities between structured objects, including molecules.

Here, we focused on a regression problem with the objective of predicting real-valued compound-target binding affinities. We used Kronecker regularized least-squares model (KronRLS) [33,49], a special variant of kernel ridge regression (KRR) which combines linear least squares with L2-norm regularization (ridge regression) and the kernel trick [50].

In KRR, given a set of N compound-protein pairs as training inputs \mathbf{x}_i 's, $i = 1, \dots, N$, and associated labels y_i 's indicating binding affinities between them, we aim to find the minimizer of the following objective function J :

$$J(f) = \sum_{i=1}^N (f(\mathbf{x}_i) - y_i)^2 + \lambda \|f\|_k^2, \quad (1)$$

where f indicates the prediction function, $f(\mathbf{x}_i)$ is the predicted binding affinity of i^{th} compound-protein pair \mathbf{x}_i , λ denotes a regularisation parameter controlling the balance between training error and model complexity ($\lambda > 0$), and $\|f\|_k$ is a norm of f on the space associated to kernel function k . In Eq (1), the first term corresponds to the training error, and the second, controlled by λ , is the penalty term that is larger for complex models that are more likely to overfit to training data but not generalize well to new instances.

According to the representer theorem [51], the prediction function that minimizes $J(f)$ can be expressed in terms of linear combination of the training examples:

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) = \boldsymbol{\alpha}^T \mathbf{k}, \quad (2)$$

where \mathbf{k} is a vector with kernel values $k(\mathbf{x}_i, \mathbf{x})$ between each training point \mathbf{x}_i and test point \mathbf{x} for which the prediction is made. The squared norm of f is therefore written as

$$\|f\|_k^2 = \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j). \quad (3)$$

A vector $\boldsymbol{\alpha}$, consisting of parameters α_i that define the solution to KRR, is found by solving the following system of linear equations

$$(\mathbf{K} + \lambda \mathbf{I}) \boldsymbol{\alpha} = \mathbf{y}, \quad (4)$$

where \mathbf{I} is the $N \times N$ identity matrix, and \mathbf{y} is the vector consisting of labels y_i . \mathbf{K} denotes $N \times N$ pairwise kernel matrix constructed for all training examples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, and thus containing similarities between all compound-protein pairs. However, the size of \mathbf{K} makes the training of the model computationally very heavy, even for moderate number of compounds and proteins.

KronRLS is a special variant of KRR, where one assumes each data point \mathbf{x}_i to consist of two separate parts, such as compound and protein, each equipped with its own kernel function, which enables to speed up the model training. Indeed, pairwise kernel \mathbf{K} is computed as the Kronecker product of compound kernel \mathbf{K}_D of size $n_D \times n_D$ and protein kernel \mathbf{K}_P of size $n_P \times n_P$ ($N = n_D \times n_P$):

$$\mathbf{K} = \mathbf{K}_D \otimes \mathbf{K}_P. \quad (5)$$

Using Eq (5), the solution to KronRLS can be calculated from:

$$\boldsymbol{\alpha} = \text{vec}(\mathbf{U}_P \mathbf{C} \mathbf{U}_D^T), \quad (6)$$

where $\text{vec}(\cdot)$ is the vectorization operator that arranges the columns of a matrix into a vector, \mathbf{U}_D and \mathbf{U}_P are orthogonal matrices with eigenvectors of drug kernel \mathbf{K}_D and protein kernel \mathbf{K}_P , respectively:

$$\mathbf{K}_D = \mathbf{U}_D \boldsymbol{\Sigma}_D \mathbf{U}_D^T, \quad (7)$$

$$\mathbf{K}_P = \mathbf{U}_P \boldsymbol{\Sigma}_P \mathbf{U}_P^T, \quad (8)$$

and \mathbf{C} is a matrix for which it holds that:

$$\text{vec}(\mathbf{C}) = (\boldsymbol{\Sigma}_D \otimes \boldsymbol{\Sigma}_P + \lambda \mathbf{I})^{-1} \text{vec}(\mathbf{U}_P^T \mathbf{Y}^T \mathbf{U}_D). \quad (9)$$

Here, $\boldsymbol{\Sigma}_D$ and $\boldsymbol{\Sigma}_P$ denote diagonal matrices containing eigenvalues of \mathbf{K}_D and \mathbf{K}_P . Label matrix \mathbf{Y} stores binding affinities between n_D drug compounds (rows) and n_P protein targets (columns). This way, we completely avoid the computation of the large pairwise kernel \mathbf{K} , and therefore significantly shorten the training time. After applying the well-known property of the Kronecker product, $(\mathbf{A} \otimes \mathbf{B})\text{vec}(\mathbf{D}) = \text{vec}(\mathbf{BDA}^T)$, the prediction for test point \mathbf{x} can be calculated as

$$f(\mathbf{x}) = (\mathbf{k}_D \otimes \mathbf{k}_P) \boldsymbol{\alpha} = (\mathbf{k}_D \otimes \mathbf{k}_P) \text{vec}(\mathbf{U}_P \mathbf{C} \mathbf{U}_D^T) = \mathbf{k}_P (\mathbf{U}_P \mathbf{C} \mathbf{U}_D^T) \mathbf{k}_D^T. \quad (10)$$

Of note, the above shortcuts work only if there are no missing values present in the label matrix \mathbf{Y} . Thus, as the pre-processing step, we imputed the missing binding affinities in \mathbf{Y} by the weighted row (compound) average. The contribution of each protein was weighted by its similarity (normalized Smith Waterman score) to the protein for which the binding affinity was missing. Such imputed values were discarded when assessing the predictive performance of the model.

The implementation of KronRLS is available at <https://github.com/aatapa/RLScore>. We tuned the regularization parameter λ of KronRLS algorithm using nested CV (Table 1).

Molecular descriptors

We computed several types of drug compound and protein target molecular descriptors in the form of kernel matrices \mathbf{K}_D and \mathbf{K}_P , respectively. The summary is presented in Fig 3C.

Drug compound space. For drug compounds, we calculated 11 fingerprint-based linear kernels and Gaussian interaction profile kernel.

Fingerprint encodes a chemical structure into a binary vector, where each bit represents the presence (1) or absence (0) of the specific substructure in the molecule. We compared two drug compounds \mathbf{d}_i and \mathbf{d}_j , represented by their fingerprints, using Tanimoto similarity score,

calculated based on the size of common substructures of the molecules:

$$\mathbf{S}_D(\mathbf{d}_i, \mathbf{d}_j) = \frac{N_{fp_i, fp_j}}{N_{fp_i} + N_{fp_j} - N_{fp_i, fp_j}}, \quad (11)$$

where fp denotes the fingerprint, N_{fp_i} is the number of 1-bits in the fingerprint fp_i of compound \mathbf{d}_i , and N_{fp_i, fp_j} indicates the number of 1-bits in fingerprints of both compounds \mathbf{d}_i and \mathbf{d}_j . The idea behind Tanimoto score is based on Jaccard's index, commonly used for comparing sample sets. Given a matrix \mathbf{S}_D of Tanimoto scores between compounds, we computed the fingerprint-based linear drug kernel as:

$$\mathbf{K}_D = \mathbf{S}_D \mathbf{S}_D^T. \quad (12)$$

We used the following 11 types of fingerprints calculated using Chemical Structure Clustering Tool of PubChem [52] (<http://pubchem.ncbi.nlm.nih.gov>) and rcdk package available for R [53].

- **D-PubChem-2D**: 881-bit fingerprint based on two-dimensional substructures defined by PubChem.
- **D-PubChem-3D**: PubChem's descriptor based on up to 10 calculated three-dimensional conformers of the compound.
- **D-PubChem-3D+**: PubChem's descriptor based on up to 10 calculated three-dimensional conformers of the compound, including additional information (acceptor count, hydrophobe count, ring count etc.).
- **D-std**: 1024-bit path-based, hashed fingerprint.
- **D-ext**: 1024-bit path-based, hashed fingerprint taking into account ring systems.
- **D-graph**: 1024-bit path-based, hashed fingerprint considering connectivity.
- **D-hybr**: 1024-bit path-based, hashed fingerprint considering hybridization states.
- **D-sp**: 1024-bit fingerprint based on the shortest paths between pairs of atoms taking into account ring systems and charges.
- **D-estate**: 79-bit fingerprint corresponding to Estate substructures described by Hall and Kier [54].
- **D-maccs**: 166-bit fingerprint based on MACCS structural keys developed by MDL Information Systems [55].
- **D-kr**: 4860-bit fingerprint defined by Klekota and Roth [56].

Furthermore, we defined a feature vector for each drug compound as its quantitative interaction profile, i.e. a feature vector \mathbf{v}_i contains binding affinities (pK_i values) between compound \mathbf{d}_i and n_p proteins present in our data set, i.e. $\mathbf{v}_i = \mathbf{Y}_i$. (i^{th} row of \mathbf{Y}). Due to data sparsity, we first imputed the missing bioactivity values in \mathbf{Y} by the weighted row average, where each row corresponds to binding affinities of a single compound. The contribution of each protein was weighted by its similarity (normalized Smith Waterman score) to the protein for which the interaction value was missing. Then, we constructed a Gaussian kernel as

follows:

$$K_{D-GIP}(\mathbf{d}_i, \mathbf{d}_j) = \exp\left(-\frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{2\sigma^2}\right), \quad (13)$$

where σ is the kernel width.

Protein target space. For kinase targets, we calculated 8 molecular descriptors, i.e. 4 linear kernels based on comparing protein amino acid sequences and three-dimensional structures, 3 generic string kernels applied to full amino acid sequences, kinase domains and ATP-binding pockets, as well as Gaussian interaction profile kernel.

We computed linear protein kernels as:

$$\mathbf{K}_p = \mathbf{S}_p \mathbf{S}_p^T, \quad (14)$$

where \mathbf{S}_p is one of the four protein-protein similarity matrices (\mathbf{S}_{P-SW} , \mathbf{S}_{P-SW+} , $\mathbf{S}_{P-3D-energy}$, $\mathbf{S}_{P-3D-sid}$), described in the following.

In the first case, we compared protein amino acid sequences using normalized version of Smith-Waterman (SW) alignment score [57]:

$$\mathbf{S}_{P-SW}(\mathbf{p}_i, \mathbf{p}_j) = \frac{SW(\mathbf{p}_i, \mathbf{p}_j)}{\sqrt{SW(\mathbf{p}_i, \mathbf{p}_i)} \sqrt{SW(\mathbf{p}_j, \mathbf{p}_j)}}, \quad (15)$$

where \mathbf{p}_i denotes i^{th} protein target, and $SW(\cdot, \cdot)$ is the original Smith-Waterman score.

Moreover, we introduced a novel molecular descriptor based on *extended target profile*. In this case, we derived features for each protein target \mathbf{p}_i present in our dataset by calculating normalized SW scores between \mathbf{p}_i and the bigger set of 20,239 human proteins (\mathbf{h}_i 's) from the UniProt database [58] (<http://www.uniprot.org/>):

$$\mathbf{S}_{P-SW+}(\mathbf{p}_i, \mathbf{h}_i) = \frac{SW(\mathbf{p}_i, \mathbf{h}_i)}{\sqrt{SW(\mathbf{p}_i, \mathbf{p}_i)} \sqrt{SW(\mathbf{h}_i, \mathbf{h}_i)}}. \quad (16)$$

Thus, given 138 protein targets in our data set, constructed \mathbf{S}_{P-SW+} matrix has the size of $138 \times 20,239$. This procedure could be thought of in terms of multiple sequence alignment, and it allows to derive global protein target features. We used BLOSUM 50 matrix in all the experiments employing SW alignments.

In order to measure pairwise similarities between three-dimensional protein structures, we used MISTRAL software, which aligns two proteins based on the minimisation of an energy function over the low-dimensional space of the relative orientations of the molecules [59]. We downloaded kinase structures from the Protein Data Bank database [60] (PDB; <http://www.rcsb.org/>). PDB files were available for 109 out of 138 kinases in our data set; we retrieved the structures of the remaining 29 molecules based on their amino acid sequence homology to proteins present in the PDB database. Then, we calculated $\mathbf{S}_{P-3D-energy}$ and $\mathbf{S}_{P-3D-sid}$ similarity matrices relating all pairs of proteins, analogously as in Eq (15) but using obtained from MISTRAL absolute value of the energy needed to align proteins or the sequence identity, respectively, instead of the $SW(\cdot, \cdot)$ scores. The sequence identity value refers to the number of aligned amino acids that are of the same chemical type.

We also compared protein targets using generic string (GS) kernel incorporating amino acid properties [44]. GS kernel compares each substring of protein sequence \mathbf{s} of size $l \leq L$

with each substring of protein sequence \mathbf{s}' having the same length:

$$\mathbf{K}_{p-GS}(\mathbf{s}, \mathbf{s}') = \sum_{l=1}^L \sum_{i=0}^{|\mathbf{s}|-l} \sum_{j=0}^{|\mathbf{s}'|-l} \exp\left(\frac{-(i-j)^2}{2\sigma_p^2}\right) \exp\left(\frac{-\|\boldsymbol{\Psi}^l(s_{i+1}, \dots, s_{i+l}) - \boldsymbol{\Psi}^l(s'_{j+1}, \dots, s'_{j+l})\|^2}{2\sigma_c^2}\right). \quad (17)$$

Here, each comparison results in a score that depends on the shifting contribution term (difference in the position of two substrings in \mathbf{s} and \mathbf{s}' , controlled by σ_p parameter) and the similarity of amino acids included in two substrings (controlled by σ_c parameter). The kernel outputs the sum of the scores from all the substring comparisons.

Vector $\boldsymbol{\Psi}^l$ is constructed as follows. Each type of amino acid a_m , $m = 1, \dots, M$, (e.g. Asparagine) has a corresponding feature vector $\boldsymbol{\Psi}(a_m)$ which defines its d properties:

$$\boldsymbol{\Psi}(a_m) = (\psi_1(a_m), \psi_2(a_m), \dots, \psi_d(a_m)). \quad (18)$$

Given a protein sequence $\mathbf{s} = s_1, s_2, \dots, s_l$, with all $s_l \in \Sigma$, where Σ is the set of all M amino acids, $\boldsymbol{\Psi}^l(\mathbf{s})$ is its encoding function which concatenates l vectors describing each amino acid the sequence \mathbf{s} is composed of:

$$\boldsymbol{\Psi}^l(\mathbf{s}) = (\boldsymbol{\Psi}(s_1), \boldsymbol{\Psi}(s_2), \dots, \boldsymbol{\Psi}(s_l)). \quad (19)$$

Furthermore, the shifting contribution term gives the GS kernel a useful property of enabling to match two amino acid subsequences even if their positions in the full protein sequences differ notably.

We used BLOSUM 50 matrix as amino acid descriptors, and we calculated three GS kernels based on comparing full protein kinase sequences, as well as sequences of kinase domains and ATP-binding pockets retrieved from the PROSITE database [61] (<http://prosite.expasy.org/>).

Finally, we calculated Gaussian protein interaction profile kernel as:

$$\mathbf{K}_{p-GIP}(\mathbf{p}_i, \mathbf{p}_j) = \exp\left(-\frac{\|\mathbf{w}_i - \mathbf{w}_j\|^2}{2\sigma^2}\right). \quad (20)$$

The formulation above is analogous to the compound interaction profile kernel defined in Eq (13), but now feature vector denoted by \mathbf{w}_i contains binding affinities between protein \mathbf{p}_i and all compounds present in our data set. All the kernel matrices are positive semidefinite and were normalized. Kernel parameters were tuned using nested cross-validation.

Model evaluation procedure

We used nested cross-validation (CV) procedure for model selection, and we assessed the predictive power with Pearson correlation and root mean squared error (RMSE) between original and predicted pK_i values.

In k -fold CV, the dataset is randomly divided into k subsamples of equal size, and the model is trained based on $k-1$ of them (training data). Then, the remaining subsample (test data) is used to assess how well the model that has been found generalizes to new instances, i.e. to calculate the predictive performance. The procedure is repeated k times, such that each subsample is used once as the test data, and the average error over the k folds gives the final estimate.

Nested CV consists of two loops, outer and inner one. In the outer CV loop, each of the k folds is kept as a test set at a time. The remaining $k-1$ training folds of the outer CV loop are further divided into training and test set of the inner CV loop. Here, the inner CV was performed during each round of the outer CV, with the aim of selecting the regularization

parameter λ of KronRLS algorithm as well as kernel parameters (S4 Table). We performed a grid search in order to select the most suitable combination of all parameters. Then, training folds of the outer CV loop were used to train the model with selected parameters, and the predictive performance was evaluated on the test set.

We applied two different CV strategies, i.e. leave-one-out nested cross-validation (LOO-CV) and leave-drug-out nested cross-validation (LDO-CV), summarized in Table 1 and S12 and S13 Figs.

We note that it is critical to discard binding affinities of compound-protein pairs belonging to the test fold of both inner and outer CV prior to computing Gaussian interaction profile kernels (KD-GIP, KP-GIP) in order to avoid significant model overfitting. Here, we used interaction profile kernels with LOO-CV, and we removed from the compound-protein interaction matrix \mathbf{Y} the whole column (row) containing the test point before computing KD-GIP (KP-GIP) kernel.

Kinase assays

Testing bioactivities predicted to fill the experimental gaps. We used Kinase Selectivity Profiling System + ADP-Glo Assay kits (Promega, TK-1 V6851, TK-2 V6853, TK-3 V6921) for performing kinase selectivity profiling at 1 μ L reaction volume on 1536-well assay plates (Corning, #3937). The reagents were prepared as indicated in Promega technical manual (TM421), and transferred to Echo Qualified 384-Well Polypropylene Microplate source plates (Labcyte, P-05525). We pre-plated five kinase inhibitors tested (cediranib, lapatinib, gefitinib, vx-745, pazopanib; all dissolved in DMSO; S5 Table) at 12 concentrations (12500, 3750, 1250, 375, 125, 37.5, 12.5, 3.75, 1.25, 0.375, 0.125, 0.0375 nM) on the 1536-well white round bottom assay plates using acoustic dispenser (Labcyte, ECHO550 Liquid handler) and DMSO transfer calibration. The inhibitors were transferred at 7.5 and 2.5 nL volumes from Echo Qualified 384-Well Low Dead Volume Microplates (Labcyte, LP-0200).

The kinase inhibitors were dissolved to 1X Kinase Buffer with 5% DMSO by transferring 200 nL of the buffer using ECHO525 Liquid handler (Labcyte). After the transfer, the assay plate was immediately sealed to minimize evaporation, and centrifuged at 872rcf/2min/RT. The dissolved kinases and 2.5X kinase buffer controls were transferred at 400 nL using ECHO525. Plates were sealed, centrifuged at 872rcf/2min/RT, mixed with plate shaker for 2 minutes and incubated 10min/RT plate upside down. The 2.5X ATP/substrate working stocks were transferred at 400 nL using ECHO525. Plates were sealed, centrifuged at 872rcf/2min/RT, mixed with plate shaker for 2 minutes, and incubated upside down (60min/RT).

For the ADP detection, 1 μ L ADP-Glo reagent was dispensed (Thermo Scientific Multidrop Combi nL). Plates were sealed, centrifuged at 872rcf/2min/RT, mixed with plate shaker for 2 minutes and incubated upside down (40 min/RT). Next, 2 μ L kinase detection reagent was dispensed with Thermo Scientific Multidrop Combi nL. Plates were sealed, centrifuged at 872rcf/2min/RT, mixed with plate shaker for 2 minutes and incubated upside down (30min/RT). Luminescence (0.5sec/well) was measured using PheraSTAR (BMG LABTECH) and 1536-plate compatible spoon.

The Kinase Selectivity Profiling System kit assays were performed at least in two individual experiments, each having two replicates.

Testing predicted targets for a kinase inhibitor tivozanib. For tivozanib's off-target testing, ABL1, RPS6KB, SLK, Aurora A, HIPK4, FYN A, FRK, as well as their substrates and buffers were purchased from SignalChem (product codes and assay concentrations available in S6 Table). Kinase assay, kinase dilution and substrate buffers were as indicated by the manufacturer (Promega). The assays were performed in the same way as with the kinases from the

Kinase Selectivity Profiling System + ADP-Glo Assay kits (see previous section). All the assays were done in three individual experiments, each having three replicates.

Analysis of the kinase assay results. IC_{50} values from all the experiments were obtained with GraphPad Prism 7 (GraphPad Software, Inc. California, USA) using *inhibitor vs. response—variable slope (four parameters)* analysis. Constrain bottom was set to *must be between zero and 100*, and constrain hill slope was set to *absolute value must be less than 1.5*. X-axis scale was set to \log_{10} . Sigmoidal dose-response curves are shown in S14 and S15 Figs.

Supporting information

S1 Fig. (a) Leave-one-out and (b) leave-drug-out cross-validation results. The prediction accuracy was assessed with root mean squared error (RMSE) between binding affinities (pK_i) from the study by Metz *et al.* and those predicted using KronRLS model with different pairs of drug (rows) and protein (columns) molecular descriptors encoded as kernel matrices (c). The lower the RMSE value, the more accurate the model prediction. Of note, Gaussian interaction profile drug kernel (KD-GIP) which resulted in the highest predictive performance under the *Bioactivity Imputation* scenario (a) has not been evaluated under the *New Drug* setup (b) because it is constructed based on the bioactivity profile of a drug, information that in practice is unavailable when predicting target interactions for a new investigational drug compound. (PDF)

S2 Fig. (a,b) Scatter plots between compound-kinase binding affinities (pK_i) measured in the Metz *et al.* study and their model predictions under the (a) *Bioactivity Imputation*, (b) *New Drug* setups, using KronRLS algorithm with the best pairs of drug and protein kernels. r indicates Pearson correlation and p -values were calculated using a Student's t distribution for a transformation of the correlation, as implemented in MATLAB Statistics Toolbox. Each point corresponds to one of 16,265 compound-kinase pairs. Most of the assays were performed at the inhibitor concentrations of 3–10,000 nM (corresponding to minimum pK_i of 5 M); however, some affinities were larger than 10,000 nM explaining the few outlier points with $pK_i < 5$ M. The higher the pK_i value, the stronger the affinity between the two molecules. Red lines mark a relatively stringent interaction threshold (7 M), distinguishing the top left corner as the region containing false positive interaction predictions, and the bottom right corner as false negative predictions. **(c,d) Receiver operating characteristic (ROC) curves constructed under the (c) *Bioactivity Imputation* and (d) *New Drug* setups using 21 different interaction threshold values (pK_i varying between 6 and 8 M with a step of 0.1 M) to binarize binding affinities measured in the Metz *et al.* study into true class labels.** Multiple values were used to study the ability of the model to discriminate interacting from non-interacting compound-kinase pairs at various interaction thresholds. The curves corresponding to the threshold of $pK_i = 7$ M, marked at the scatter plots (a) and (b), are plotted with the darkest red colour. AUC indicates the area under the ROC curve; the closer AUC is to 1, the more accurate the model prediction. (PDF)

S3 Fig. Distribution of 152 drug-wise Pearson correlation values between compound-kinase binding affinities (pK_i) measured in the Metz *et al.* study and their model predictions under the *New Drug* scenario. The predictions were made using KronRLS algorithm with the best pair of drug and protein kernels (KD-sp and KP-GS) under the leave-drug-out cross-validation. 125 out of 152 correlations are statistically significant ($p < 0.0001$), and 58 correlation values are greater than 0.75. (PDF)

S4 Fig. Scatter plots between the measured compound-kinase binding affinities from Metz *et al.* study and their model predictions obtained using KronRLS with KD-GIP drug kernel and KP-GS-domain protein kernel. r indicates Pearson correlation and p -values were calculated using a Student's t distribution for a transformation of the correlation, as implemented in MATLAB Statistics Toolbox. Each plot corresponds to a single compound selected for the experimental validation, and each point represents its bioactivity against a single kinase. Points marked with red stars and orange dots are predictions for the unmeasured in Metz *et al.* study compound-kinase pairs; thus, they are placed on the y -axis. Red star-shapes and green triangular points indicate interactions included in our experimental validation. However, the green points constitute merely an experimental control as those binding affinities are reported in Metz *et al.* data set. The higher the pK_i value, the stronger the affinity between the two molecules. Red lines mark a pre-defined interaction threshold (7 M), distinguishing top left corner as the region containing false positive (FPs) interaction predictions, and bottom right corner as false negative (FN) predictions. During the selection of the molecules for the experimental validation, we aimed at minimizing the number of FPs and FNs. For instance, in case of cediranib (a), there are no FPs, meaning that the predicted interaction, indicated by the red star, is likely a true one. The percentage of missing values indicates unmeasured binding affinities of a given compound against a set of 138 kinases in the Metz *et al.* data set.

(PDF)

S5 Fig. The comparison between model-predicted and experimentally-measured in different assays bioactivities of 100 compound-kinase pairs included in our experimental validation. (a,c) Technical variability between two experimental kinase assays. Scatter plots between (a) 82 pK_i values measured in Metz *et al.* study and pIC_{50} values from our experimental assay; (c) 95 pK_d values from Davis *et al.* study and pIC_{50} values from our experimental assay; (b) 73 *in silico*-predicted and measured in Metz *et al.* study pK_i values, excluding compound-kinase pairs blinded in the model training (marked with orange colour in Fig 4A and S2 Table); (d) 95 *in silico*-predicted pK_i values and pK_d readouts from Davis *et al.* study. The values are detailed in S2 Table.

(PDF)

S6 Fig. The comparison between model-predicted (based on the data from Metz *et al.* study) and experimentally-measured (in the study by Davis *et al.*) compound-kinase bioactivities. (a) The *Bioactivity Imputation* scenario. The KronRLS model was trained using Metz *et al.* dataset together with the best-performing (under the *Bioactivity Imputation* scenario) drug interaction profile kernel (KD-GIP) and kinase domain-based generic string protein kernel (KP-GS-domain, Fig 3A). The model was then used to predict binding affinities between 2,662 compound-kinase pairs overlapping between Metz *et al.* and Davis *et al.* datasets. (b) The *New Target* scenario. The KronRLS algorithm was trained using Metz *et al.* dataset together with the best-performing (under the *New Target* scenario) PubChem's fingerprint-based drug kernel (KD-PubChem-2D) and extended target profile protein kernel built upon Smith-Waterman amino acid sequence comparisons (KP-SW+, S7 Fig). The model was then used to predict the binding affinities between 152 drugs from Metz *et al.* study and 244 wild-type kinases present in Davis *et al.* but not Metz *et al.* dataset. The predictive performance was evaluated based on 5,368 binding affinities of 22 drugs overlapping between the two datasets.

(PDF)

S7 Fig. (a) Leave-target-out cross-validation results. The prediction accuracy was evaluated with Pearson correlation (r) between binding affinities (pK_i) from the study by Metz *et al.* and those predicted using KronRLS algorithm with different pairs of drug (rows) and protein

(columns) molecular descriptors encoded as kernel matrices (b).
(PDF)

S8 Fig. Kinase dendrogram created by hierarchical clustering of 138 kinases from Metz *et al.* study based on their bioactivity data (pK_i). The bar length is proportional to the number of compound interactions for each kinase ($pK_i \geq 7$ M). On-targets of tivozanib are marked with red lines (*FLT1*, *FLT4*, *KDR*), fedratinib—green lines (*JAK2*, *JAK3*, *TYK2*), vx11e and ulixertinib (a compound derived from vx11e)—blue lines (vx11e: *CDK2*, *ERK2*; ulixertinib: *ERK2*).
(PDF)

S9 Fig. Interaction map between 152 compounds (rows) and 138 kinases (columns) profiled in the study of Metz *et al.* The white cells represent unmeasured binding affinities. The higher the pK_i value, the stronger the affinity between the compound and kinase.
(PDF)

S10 Fig. Distribution of 16,265 compound-kinase binding affinities measured in the study of Metz *et al.*
(PDF)

S11 Fig. Kinome map of 138 kinases used in our work. Figure was created with KinMap (<http://kinhub.org/kinmap>).
(PDF)

S12 Fig. Schematic illustration of the nested leave-one-out cross-validation (LOO-CV) procedure, consisting of the inner loop for model selection, and the outer loop for model performance estimation. Single round of the outer CV is shown, where one compound-protein pair is removed from the training data, and used as a test fold. The inner leave-one-out CV is run for each model parameter combination (grid search) during every round of the outer CV. The combination resulting in the lowest root mean squared error between the original and predicted binding affinities is selected and used in the model training in the outer CV loop. We used LOO-CV to tune the model parameters and assess its performance under the *Bioactivity Imputation* scenario (Fig 2A).
(PDF)

S13 Fig. Schematic illustration of the nested leave-drug-out cross-validation (LDO-CV) procedure consisting of the inner loop for model selection, and the outer loop for model performance estimation. Single round of the outer CV is shown, where all binding affinities of selected compound are removed from the training data, and used as a test fold. The inner leave-drug-out CV is run for each model parameter combination (grid search) during every round of the outer CV. In the inner CV, five compounds are removed at a time (at random). The model parameter combination resulting in the lowest root mean squared error between the original and predicted binding affinities is selected and used in the model training in the outer CV loop. Note that also the rows and columns corresponding to the compounds included in the test fold are removed from the drug kernel matrix before model training. We used LDO-CV to tune the model parameters and assess its performance under the *New Drug* scenario (Fig 2B).
(PDF)

S14 Fig. Results of our kinase assay for testing bioactivities predicted to fill the experimental gaps in the large-scale kinase inhibitor target profiling study by Metz *et al.*; examples of drug response curves obtained as described in Materials and Methods section of the main

paper. Corresponding pIC_{50} values are summarized in [S2 Table](#).
(PDF)

S15 Fig. Results of our kinase assay for testing predicted target interactions for a new investigational kinase inhibitor tivozanib; drug response curves obtained as described in Materials and Methods section of the main paper. Corresponding pIC_{50} values are summarized in [S3 Table](#).
(PDF)

S16 Fig. Visualisation of the binding between tivozanib and ABL1 (PDB code: 2e2b). The docking was performed with Rosetta (<https://www.rosettacommons.org/>) and the figure was created using UCSF cHimera (<https://www.cgl.ucsf.edu/chimera/>). A radius for docking was set to 5 Å around the centre of the ATP-binding site.
(PNG)

S1 Table. Possible drug-protein interaction prediction scenarios. In this work, we focused on two most practical ones, namely the *Bioactivity Imputation* ([Fig 2A](#)) and the *New Drug* ([Fig 2B](#)) scenarios. Additionally, we included the results under the *New Target* setup ([Fig 2C](#)) in [S7 Fig](#). (d_x, p_x) denotes the query drug-protein pair, the binding affinity of which one aims to predict.
(PDF)

S2 Table. Results of the experimental validation of bioactivities predicted to fill the gaps in the study of Metz *et al.*
(XLSX)

S3 Table. Predicted target profile of tivozanib and the results of experimental validation.
(XLSX)

S4 Table. Model parameters tested. In case of Gaussian kernels (K_{D-GIP} , K_{P-GIP}), the values for kernel width parameter σ were selected by computing pairwise distances between all data points and taking 0.1, 0.5 and 0.9 quantiles.
(PDF)

S5 Table. Kinase inhibitors used in our experimental assays.
(PDF)

S6 Table. Kinases, substrates, and their concentrations used in tivozanib's off-target testing.
(PDF)

Acknowledgments

The authors would like to thank Dr. Lari Lehtiö for careful reading and expert comments on the manuscript, Dr. Simon Anders for his statistical advice, as well as Mohamed Jabri for performing the molecular docking of tivozanib to ABL1.

Author Contributions

Conceptualization: Anna Cichonska, Balaguru Ravikumar, Krister Wennerberg, Juho Rousu, Tero Aittokallio.

Data curation: Anna Cichonska.

Formal analysis: Anna Cichonska.

Funding acquisition: Anna Cichonska, Balaguru Ravikumar, Krister Wennerberg, Juho Rousu, Tero Aittokallio.

Investigation: Anna Cichonska, Elina Parri, Sanna Timonen.

Methodology: Anna Cichonska, Elina Parri, Sanna Timonen, Tapio Pahikkala, Antti Airola.

Project administration: Anna Cichonska, Tero Aittokallio.

Resources: Krister Wennerberg, Juho Rousu, Tero Aittokallio.

Software: Anna Cichonska, Tapio Pahikkala, Antti Airola.

Supervision: Krister Wennerberg, Juho Rousu, Tero Aittokallio.

Validation: Anna Cichonska, Sanna Timonen, Krister Wennerberg, Juho Rousu, Tero Aittokallio.

Visualization: Anna Cichonska, Balaguru Ravikumar.

Writing – original draft: Anna Cichonska, Tero Aittokallio.

Writing – review & editing: Anna Cichonska, Balaguru Ravikumar, Elina Parri, Tapio Pahikkala, Antti Airola, Krister Wennerberg, Juho Rousu, Tero Aittokallio.

References

1. Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer*. 2010; 10:130–7. <https://doi.org/10.1038/nrc2787> PMID: 20094047
2. Hu Y, Furtmann N, Bajorath J. Current compound coverage of the kinome: miniperspective. *J Med Chem*. 2014; 58:30–40. <https://doi.org/10.1021/jm5008159> PMID: 25051177
3. Metz JT, Johnson EF, Soni NB, Merta PJ, Kifle L, Hajduk PJ. Navigating the kinome. *Nat Chem Biol*. 2011; 7:200–2. <https://doi.org/10.1038/nchembio.530> PMID: 21336281
4. Savitski MM, Reinhard FB, Franken H, Werner T, Savitski MF, Eberhard D, et al. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science*. 2014; 346:1255784. <https://doi.org/10.1126/science.1255784> PMID: 25278616
5. Reymond JL, Awale M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem Neurosci*. 2012; 3:649–57. <https://doi.org/10.1021/cn3000422> PMID: 23019491
6. Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug-target interactions: a brief review. *Brief Bioinform*. 2014; 15:734–47. <https://doi.org/10.1093/bib/bbt056> PMID: 23933754
7. Yamanishi Y. Chemogenomic approaches to infer drug-target interaction networks. In *Data Mining for Systems Biology: Methods and Protocols*. 2013; 939:97–113.
8. Mousavian Z, Masoudi-Nejad A. Drug-target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin Drug Metab Toxicol*. 2014; 10:1273–87. <https://doi.org/10.1517/17425255.2014.950222> PMID: 25112457
9. Cichonska A, Rousu J, Aittokallio T. Identification of drug candidates and repurposing opportunities through compound-target interaction networks. *Expert Opin Drug Discov*. 2015; 10:1333–45. <https://doi.org/10.1517/17460441.2015.1096926> PMID: 26429153
10. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, et al. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform*. 2016; 17:696–712. <https://doi.org/10.1093/bib/bbv066> PMID: 26283676
11. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: where have you been? Where are you going to? *J Med Chem*. 2014; 57:4977–5010. <https://doi.org/10.1021/jm4004285> PMID: 24351051
12. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform*. 2010; 29:476–88. <https://doi.org/10.1002/minf.201000061> PMID: 27463326
13. Li Q, Lai L. Prediction of potential drug targets based on simple sequence properties. *BMC Bioinformatics*. 2007; 8:353. <https://doi.org/10.1186/1471-2105-8-353> PMID: 17883836

14. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*. 2008; 24:i232–40. <https://doi.org/10.1093/bioinformatics/btn162> PMID: 18586719
15. Bosc N, Wroblowski B, Meyer C, Bonnet P. Prediction of protein kinase-ligand interactions through 2.5D kinochemometrics. 2017; 57: 93–101. <https://doi.org/10.1021/acs.jcim.6b00520> PMID: 27983837
16. Merget B, Turk S, Eid S, Rippmann F, Fulle, S. Profiling prediction of kinase inhibitors: toward the virtual assay. *J Med Chem*. 2017; 60:474–85. <https://doi.org/10.1021/acs.jmedchem.6b01611> PMID: 27966949
17. Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, et al. Deep learning-based drug-target interaction prediction. *J Proteome Res*. 2017; 16:1401–9. <https://doi.org/10.1021/acs.jproteome.6b00618> PMID: 28264154
18. Zhang X, Li L, Ng MK, Zhang S. Drug-target interaction prediction by integrating multiview network data. *Comput Biol Chem*. 2017; In Press.
19. Christmann-Franck S, van Westen GJ, Papadatos G, Escudie FB, Roberts A, Overington JP, et al. Unprecedentedly large-scale kinase inhibitor set enabling the accurate prediction of compound-kinase activities: a way toward selective promiscuity by design? *J Chem Inf Model*. 2016; 56:1654–75. <https://doi.org/10.1021/acs.jcim.6b00122> PMID: 27482722
20. Lim H, Poleksic A, Yao Y, Tong H, He D, Zhuang L, et al. Large-scale off-target identification using fast and accurate dual regularized one-class collaborative filtering and its application to drug repurposing. *PLoS Comput Biol*. 2016; 12:e1005135. <https://doi.org/10.1371/journal.pcbi.1005135> PMID: 27716836
21. Coelho ED, Arrais JP, Oliveira JL. Computational discovery of putative leads for drug repositioning through drug-target interaction prediction. *PLoS Comput Biol*. 2016; 12:e1005219. <https://doi.org/10.1371/journal.pcbi.1005219> PMID: 27893735
22. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics*. 2011; 27:3036–43. <https://doi.org/10.1093/bioinformatics/btr500> PMID: 21893517
23. van Laarhoven T, Marchiori E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS One*. 2013; 8:e66952. <https://doi.org/10.1371/journal.pone.0066952> PMID: 23840562
24. Alaimo S, Pulvirenti A, Giugno R, Ferro A. Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics*. 2013; 29:2004–8. <https://doi.org/10.1093/bioinformatics/btt307> PMID: 23720490
25. Gönen M. Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. *Bioinformatics*. 2012; 28:2304–10. <https://doi.org/10.1093/bioinformatics/bts360> PMID: 22730431
26. Liu Y, Wu M, Miao C, Zhao P, Li XL. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput Biol*. 2016; 12:e1004760. <https://doi.org/10.1371/journal.pcbi.1004760> PMID: 26872142
27. Wang Y, Zeng J. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics*. 2013; 29:i126–34. <https://doi.org/10.1093/bioinformatics/btt234> PMID: 23812976
28. Perlman L, Gottlieb A, Atias N, Rupp E, Sharan R. Combining drug and gene similarity measures for drug-target elucidation. *J Comput Biol*. 2011; 18:133–45. <https://doi.org/10.1089/cmb.2010.0213> PMID: 21314453
29. Yuan Q, Gao J, Wu D, Zhang S, Mamitsuka H, Zhu S. DrugE-Rank: improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics*. 2016; 32:i18–27. <https://doi.org/10.1093/bioinformatics/btw244> PMID: 27307615
30. Bora A, Avram S, Ciucanu I, Raica M, Avram S. Predictive models for fast and effective profiling of kinase inhibitors. *J Chem Inf Model*. 2016; 56:895–905. <https://doi.org/10.1021/acs.jcim.5b00646> PMID: 27064988
31. Li ZC, Huang MH, Zhong WQ, Liu ZQ, Xie Y, Dai Z, et al. Identification of drug-target interaction from interactome network with “guilt-by-association” principle and topology features. *Bioinformatics*. 2016; 32:1057–64. <https://doi.org/10.1093/bioinformatics/btv695> PMID: 26614126
32. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, et al. Toward more realistic drug-target interaction predictions. *Brief Bioinform*. 2015; 16:325–37. <https://doi.org/10.1093/bib/bbu010> PMID: 24723570
33. Pahikkala T, Airola A, Stock M, De Baets B, Waegeman W. Efficient regularized least-squares algorithms for conditional ranking on relational data. *Mach Learn*. 2013; 93:321–56.

34. Ceroni A, Costa F, Frasconi P. Classification of small molecules by two-and three-dimensional decomposition kernels. *Bioinformatics*. 2007; 23:2038–45. <https://doi.org/10.1093/bioinformatics/btm298> PMID: 17550912
35. Marcou G, Horvath D, Varnek A. Kernel target alignment parameter: a new modelability measure for regression tasks. *J Chem Inf Model*. 2015; 56:6–11. <https://doi.org/10.1021/acs.jcim.5b00539> PMID: 26673976
36. Cer RZ, Mudunuri U, Stephens R, Lebeda FJ. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res*. 2009; 37:W441–5. <https://doi.org/10.1093/nar/gkp253> PMID: 19395593
37. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model*. 2014; 54:735–43. <https://doi.org/10.1021/ci400709d> PMID: 24521231
38. Kallioikoski T, Kramer C, Vulpetti A, Gedeck P. Comparability of mixed IC50 data—a statistical analysis. *PLoS One*. 2013; 8:e61007. <https://doi.org/10.1371/journal.pone.0061007> PMID: 23613770
39. Knight ZA, Shokat KM. Features of selective kinase inhibitors. *Chem Biol*. 2005; 12:621–37. <https://doi.org/10.1016/j.chembiol.2005.04.011> PMID: 15975507
40. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2011; 29:1046–51. <https://doi.org/10.1038/nbt.1990> PMID: 22037378
41. Jamil MO, Hathaway A, Mehta A. Tivozanib: status of development. *Curr Oncol Rep*. 2015; 17:1–7. <https://doi.org/10.1007/s11912-014-0425-x>
42. Gao Y, Davies SP, Augustin M, Woodward A, Patel UA, Kovelman R, et al. A broad activity screen in support of a chemogenomic map for kinase signalling research and drug discovery. *Biochem J*. 2013; 451:313–28. <https://doi.org/10.1042/BJ20121418> PMID: 23398362
43. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res*. 2014; 42:1083–90.
44. Giguère S, Marchand M, Laviolette F, Drouin A, Corbeil J. Learning a peptide-protein binding affinity predictor with kernel ridge regression. *BMC Bioinformatics*. 2013; 14:82. <https://doi.org/10.1186/1471-2105-14-82> PMID: 23497081
45. Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol*. 2014; 32:1202–12. <https://doi.org/10.1038/nbt.2877> PMID: 24880487
46. Brouard C, Shen H, Dührkop K, d'Alché-Buc F, Böcker S, Rousu J. Fast metabolite identification with input output kernel regression. *Bioinformatics*. 2016; 32:i28–36. <https://doi.org/10.1093/bioinformatics/btw246> PMID: 27307628
47. Nascimento AC, Prudêncio RB, Costa IG. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinformatics*. 2016; 17:46. <https://doi.org/10.1186/s12859-016-0890-3> PMID: 26801218
48. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol*. 2011; 29:1039–45. <https://doi.org/10.1038/nbt.2017> PMID: 22037377
49. Pahikkala T, Airola A. RLScore: regularized least-squares learners. *J Mach Learn Res*. 2016; 17:1–5.
50. Saunders C, Gammerman A, Vovk V. Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*; 1998 Jul 24–27; Madison, Wisconsin, USA. Morgan Kaufmann Publishers; 1998.
51. Kimeldorf G, Wahba G. Some results on Tchebycheffian spline functions. *J Math Anal Appl*. 1971; 33:82–95.
52. Kim S, Thiessen PA, Bolton EE, Bryant SH. PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. *Nucleic Acids Res*. 2015; 43:W605–11. <https://doi.org/10.1093/nar/gkv396> PMID: 25934803
53. Guha R. Chemical informatics functionality in R. *J Stat Softw*. 2007; 18:1–16.
54. Hall LH, Kier LB. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J Chem Inf Comput Sci*. 1995; 35:1039–45.
55. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci*. 2002; 42:1273–80. PMID: 12444722
56. Klekota J, Roth FP. Chemical substructures that enrich for biological activity. *Bioinformatics*. 2008; 24:2518–25. <https://doi.org/10.1093/bioinformatics/btn479> PMID: 18784118

57. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981; 147:195–7. PMID: [7265238](#)
58. Consortium UniProt. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015; 43:D204–12. <https://doi.org/10.1093/nar/gku989> PMID: [25348405](#)
59. Micheletti C, Orland H. MISTRAL: a tool for energy-based multiple structural alignment of proteins. *Bioinformatics.* 2009; 25:2663–9. <https://doi.org/10.1093/bioinformatics/btp506> PMID: [19692555](#)
60. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000; 28:235–42. PMID: [10592235](#)
61. Sigrist CJA, de Castro E, Cerutti L, Cuče BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* 2012; 41:D344–7. <https://doi.org/10.1093/nar/gks1067> PMID: [23161676](#)